

Dansk webarkivering

– en "afspejling af den menneskelige tilværelse i det moderne samfund" og nutidens bidrag til den danske kulturarv

Af Lennart Björneborn

Vi kan i dag se bevarede eksemplarer af de første trykte bøger for over 500 år siden. Vi kan også se, hvordan de første aviser og de første videnskabelige tidsskrifter så ud. Vi kan ydermere høre de første radioprogrammer og se de første tv-udsendelser. Men vi kan i dag ikke se, hvordan de første websider så ud for kun godt 10 år siden. De er nemlig desværre ikke blevet indsamlet og bevaret for eftertiden.

Lennart Björneborn
Institut for Informationsstudier
Danmarks Biblioteksskole
lb@db.dk



Danmarks Biblioteksskole arrangerede den 24. februar i år en temadag om "Dansk webarkivering - indsamling og bevaring af digital kulturarv". Temadagen var aktualiseret af den nye pligtafleveringslov, der træder i kraft 1. juli i år og som lægger op til, at Danmark nu får et nationalt webarkiv, der indsamler og bevarer danske websider for eftertiden.

Jeg vil her give nogle ganske korte indtryk fra temadagen. Overheads og dias fra alle oplæggene kan hentes på temadagens hjemmeside¹.

Temadagen startede med, at Niels Ole Finnemann, leder af Center for Internetforskning², Århus Universitet, redegjorde for, hvordan internettet bidrager til vores kulturarv. Ifølge Finnemann giver internettet "den mest omfattende, alsidige og dækkende afspejling af den menneskelige tilværelse i de moderne samfund" og udgør derfor "et af nutidens originale bidrag til det danske og andre moderne samfunds kulturarv". I denne forbindelse kan nævnes, at P1 på Danmarks Radio sidste år sendte et radioprogram om 'Webhøstning af den digitale kulturarv', hvor bl.a. Niels Ole Finnemann blev interviewet. Lydfilen med radioprogrammet er tilgængelig på DRs hjemmeside³.

I mit eget indlæg på temadagen beskrev jeg, hvordan de fleste websider fra WWWs barndom i starten af 1990'erne desværre er gået tabt, fordi de ikke er blevet arkiveret, som nævnt ovenfor. Men heldigvis har det amerikansk-baserede private initiativ, Internet Archive, indsamlet over 30 milliarder websider på verdensplan siden 1996, inkl. mange danske websider. Via deres Wayback Machine⁴ kan man nemt indtaste webadresser på de sider fra webbens 'oldtid', man gerne vil se. I mit indlæg gav jeg nogle eksempler fra mit eget nyligt afsluttede ph.d.-projekt om linkstrukturer på

WWW⁵, hvor jeg bl.a. kunne bruge Internet Archive som et web-'arkæologisk' redskab til at genfinde 'gamle' websider og links.

På temadagen deltog også Allan Arvidson fra Kungliga Biblioteket i Stockholm, der står for det svenske Kulturarw3-projekt⁶, der siden 1996 har indsamlet millioner af svenske websider. Arvidson beklagede, at der desværre kun er ganske få brugere af det svenske webarkiv, da der kun er adgang til det fra publikumsskinner på selve KB i Stockholm. Adgangsbe- grænsningerne skyldes hensyn til de potentielt følsomme personoplysninger, der kan være på arkiverede websider.

Harald von Hielmcrone fra Statsbiblioteket og Birgit Nordmark Henriksen fra det Kongelige Bibliotek fortalte om den nye pligtafleveringslov og hvad den kan betyde for, hvordan webarkivering gennemføres i Danmark. Statsbiblioteket og Det Kongelige Bibliotek opbygger sammen det danske webarkiv på baggrund af erfaringer fra pilotprojektet netarkivet.dk⁷.

Der skal indsamles materiale til det nye webarkiv på tre forskellige måder:

- 1) Tværsnitshøstning, der 4 gange årligt indsamler materiale i bredden fra alle de ca. 200.000 registrerede websteder, der er aktive i dk-domænet samt identificerede 'danica'-sider uden for dette domæne.
- 2) Selektiv høstning, der mere dybtgående og fuldstændigt indsamler fra ca. 80 udvalgte websteder, som fx folketinget, politiske partier, landsdækkende aviser og tv-stationer.

- 3) Begivenhedshøstning, der indsamler webmateriale både i bredden og dybden fra et par særlige begivenheder hvert år, som fx kommunevalg, folketingsvalg, EU-afstemning, OL eller et kongeligt bryllup.

Birgit Nordmark Henriksen omtalte de tekniske udfordringer, der ligger i at langtids-sikre og tilgængeliggøre de indsamlede digitale materialer, så man også langt ud i fremtiden vil kunne fremvise materialernes oprindelige organisering, grafiske layout, hypertextfunktionalitet, osv.

Danmark udveksler på dette område erfaringer med Internet Archive samt en række lande, heriblandt USA (Library of Congress), Australien, Frankrig og de nordiske lande, i et internationalt samarbejde, International Internet Preservation Consortium⁸.

Birgit Nordmark Henriksen har i april 2005 skrevet en artikel, der giver et meget godt overblik over webarkivering⁹ internationalt og i Danmark.

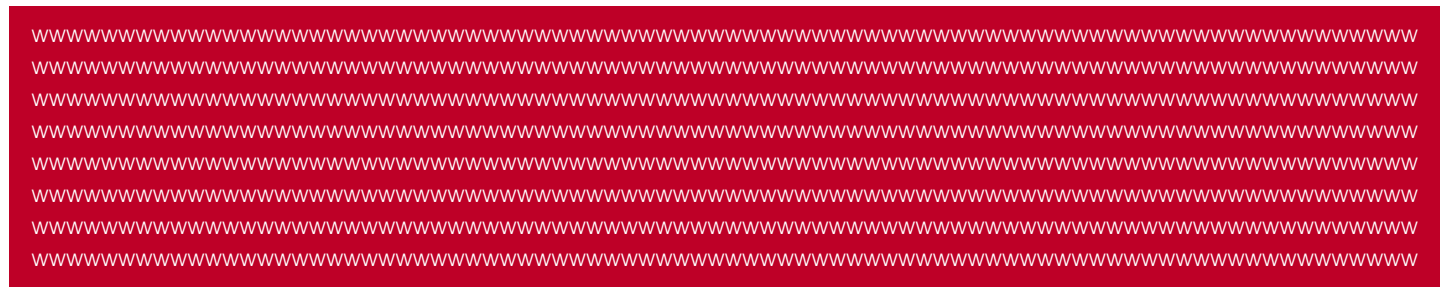
Pligtafleveringsloven

Med den ny pligtafleveringslov, der som nævnt træder i kraft 1. juli i år, bliver det webarkivets opgave at lokalisere og indsamle danske webmaterialer. I den forbindelse er det vigtigt, at webstederne (herunder også forskningsbibliotekerne og de højere læreanstalter) placerer og tilgængeliggør deres offentliggjorte webdokumenter tilpas nær på 'overfladen' af deres websteder, så webarkivets crawlere kan nå disse dokumenter via interne links på webstedet.

WORLD WIDE WEB

www://den mest omfattende, alsidige og dækkende afspejling af den menneskelige tilværelse i de moderne samfund.

Niels Ole Finnemann



Potentielt følsomme personoplysninger på arkiverede websider giver problemer med adgang til det danske webarkiv, ligesom det er tilfældet i Sverige, som nævnt ovenfor. Krav fra Datatilsynet gør, at kun forskere med godkendte forskningsprojekter kan få adgang til arkivet. Interessant nok kunne Hielmcrone oplyse, at de nuværende almindelige søgemaskiner faktisk ikke opfylder Datatilsynets krav om beskyttelse af potentielt følsomme personoplysninger på websider. Eksempler på dette kunne ifølge Hielmcrone være websider, hvor personer hænger en eksklusiv ud, eller en SM-forening, der lægger deres medlemsliste ud på webben, selv om ikke alle medlemmer nødvendigvis ønsker dette.

Alt andet lige er ovenstående eksempler – sammen med alle de millioner af andre websider, der vil blive arkiveret fremover – en ”*afspejling af den menneskelige tilværelse i det moderne samfund*” med Finnemanns ord. Så fremtidens web-’arkæologer’ kan sikkert bruge webarkiverne til at grave sig frem til nok så alsidige og interessante detaljer om nutidens levevis og normer!

NOTER

- 1 www.db.dk/kon/temadag/f2005/danskwebarkiv.htm
- 2 www.cfi.imv.au.dk
- 3 www.dr.dk/p1/sommer2004/kulturarv.asp?action=showarticle&id=32142
- 4 www.archive.org/
- 5 www.db.dk/lb/
- 6 www.kb.se/kw3/
- 7 www.netarkivet.dk
- 8 www.netpreserve.org
- 9 <http://netarkivet.dk/website/publications/webarkivering-webarchiving.pdf>



Temadagen var aktualiseret af den nye pligtafleveringslov, der træder i kraft 1. juli i år og som lægger op til, at Danmark nu får et nationalt webarkiv, der indsamler og bevarer danske websider for eftertiden.